City Research Online

City, University of London Institutional Repository

Citation: Andrienko, G. and Andrienko, N. (2013). Spatio-temporal analysis of flows in CDC 2013 data. Paper presented at the Understanding Urban Cycling: A Data Challenge, 14 Nov 2013, Leuven, Belgium.

This is the unspecified version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/2905/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Spatio-temporal analysis of flows in CDC 2013 data

Gennady Andrienko
Fraunhofer Institute IAIS
Schloss Birlinghoven
53757 Sankt Augustin
Germany
gennady.andrienko@iais.fraunhofer.de

Natalia Andrienko
Fraunhofer Institute IAIS
Schloss Birlinghoven
53757 Sankt Augustin
Germany
natalia.andrienko@iais.fraunhofer.de

Abstract

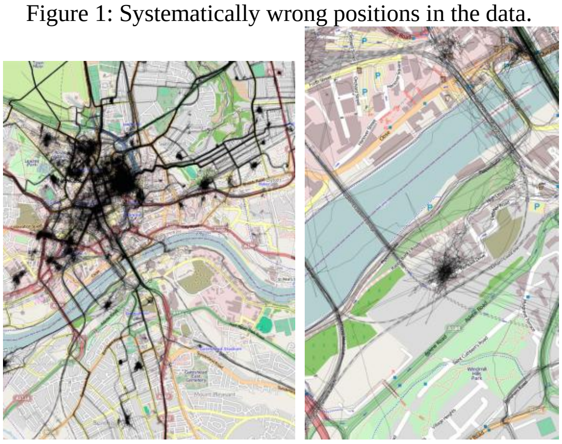
We describe analysis of flows in the CDC2013 bicycles data set.
Keywords: movement data, flows, clustering.

1 Data preprocessing

In this study, we analyzed the CDC2013 bicycles data set[1] aiming at characterizing streets according to their suitability for bicycle traffic. The available data included 2,403,904 time-stamped GPS positions of 78 distinct travellers recorded from the 28th of June, 2011 till the 1st of December, 2011. Only two positions have been recorded before the 7th of October, 2011; they have been discarded.

We performed pre-processing consisting of (1) eliminating duplicates, that is, records having the same person ID and time stamp, (2) eliminating stationary points characterized by speed lower than 2km/h, (3) dividing the tracks into daily trajectories by breaking them at 3AM, (4) dividing the daily trajectories by stops for at least 30 minutes or spatial gaps of more than 1km between positions, (5) eliminating short mostly stationary trajectories consisting of less than 5 points or having duration below 5 minutes or localized within an area with less than 50 metres radius. As a result, we obtained 1,584 trajectories representing trips.

Figure 1: Systematically wrong positions in the data.

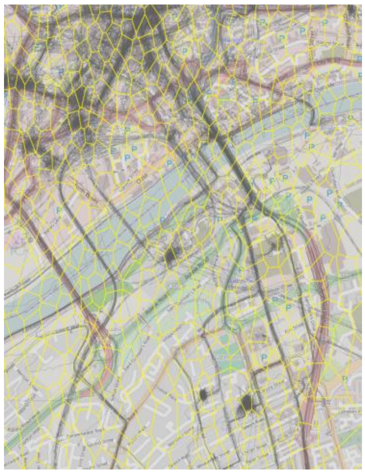


By visual inspection of maps (Figure 1), we found that many trajectories consist of systematically wrong positions. We partly improved the data by removing trajectory segments characterized by very high speed, tortuosity, or sinuosity, as well as stops expressed in the data as random changes of positions around one place. Despite our efforts, the data remained far from perfect (probably, due to misconfiguration of hardware and/or software during the data collection campaign), prohibiting many kinds of potentially interesting analyses. Moreover, we found that even major streets were used by quite small number of bicyclists, and that observations of most individuals were shorter than one week. These facts further restrict possible types of analysis [1].

2 Data processing

For data aggregation, we produced a Voronoi tessellation of the territory reflecting the spatial distribution of the recorded positions [2]. The method extracted characteristic points of trajectories and clustered them in space, using a chosen radius of 100m for splitting dense clusters covering large areas.

Figure 2: Voronoi tessellation of the territory.



[1] https://sites.google.com/site/cdc2013workshop/

This procedure resulted in 14,033 polygons (Figure 2). It can be observed that the tessellation correctly reflects the street network.

3 Data analysis

Based on the computed tessellation, we calculated flows between the polygons (Figure 3) and hourly time series (Figure 4) of the transitions. Aggregated time series plot shows statistical distributions (deciles) of flow magnitude values over time (Figure 4 bottom).

Figure 3: Map fragment presents a part of 26,094 connections; the maximal flow magnitude is 127.

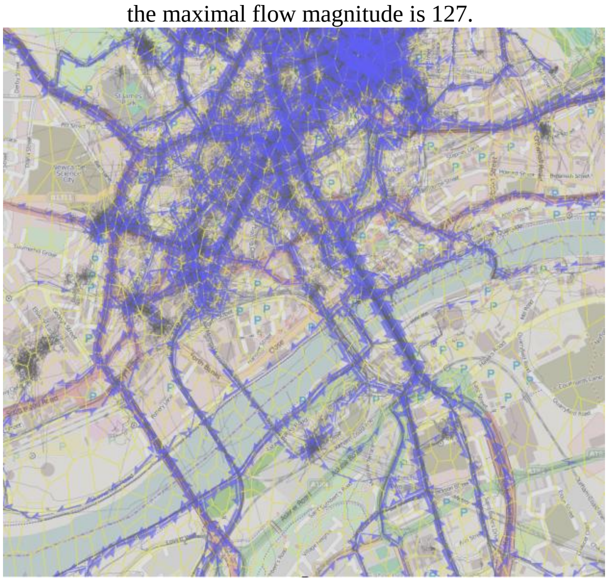
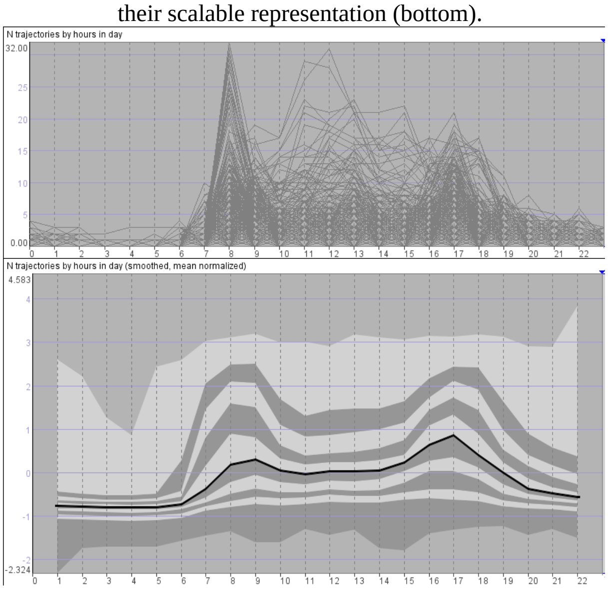


Figure 4: Hourly time series of flow magnitudes (top) and their scalable representation (bottom).



Time series have been used for clustering the flows by similarity by kMeans. Spatial footprints and temporal profiles of 5 major clusters are shown in Figure 5.

Figure 5: Spatial and temporal profiles of 5 clusters.

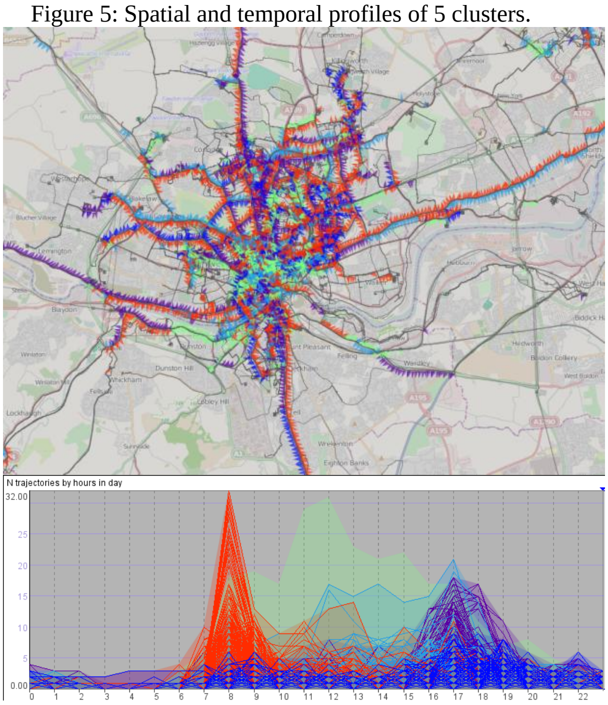


Figure 6: Cluster 5: mostly mid-day flows

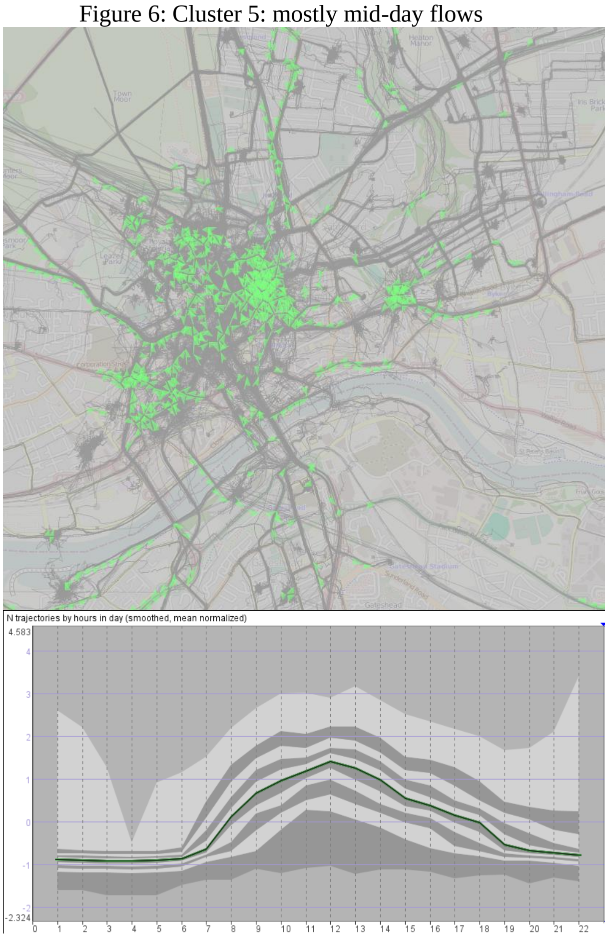


Figure 7: Cluster 3: morning activities.

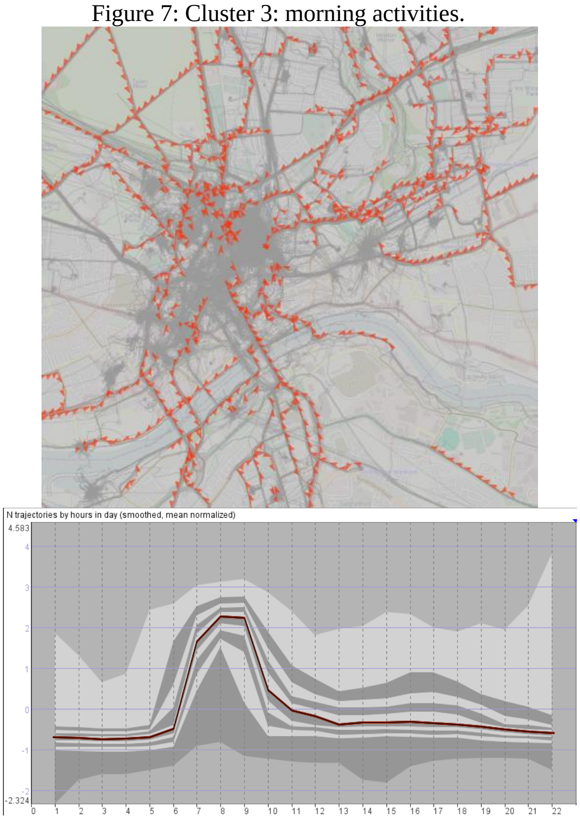


Figure 8: Cluster 1: evening activities.

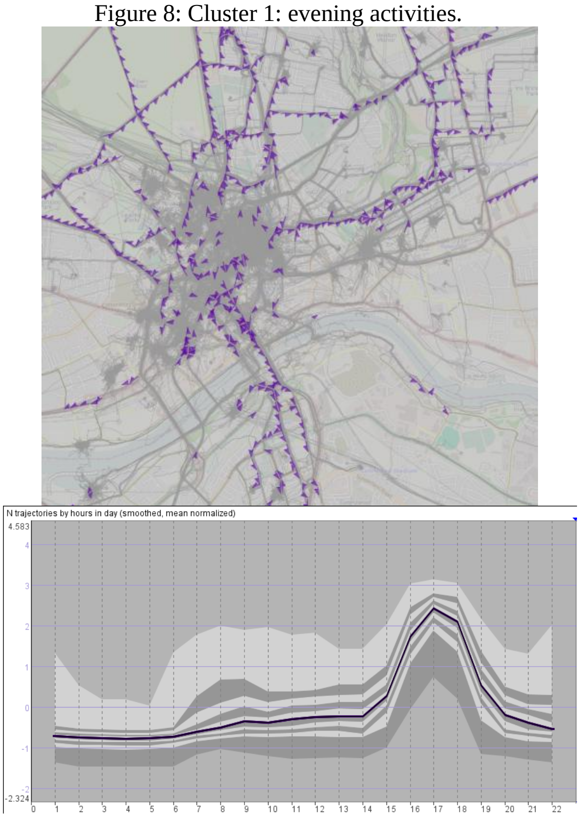


Figure 9: Cluster 2: late afternoon activities.

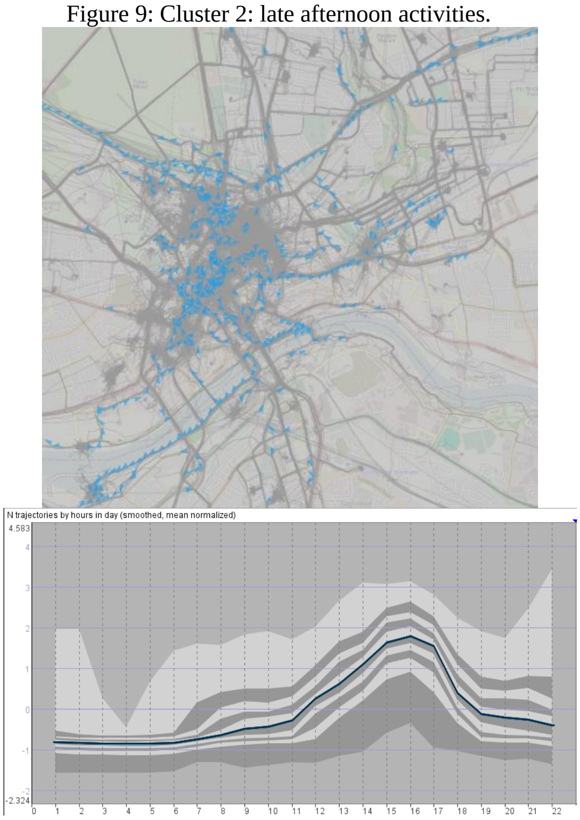
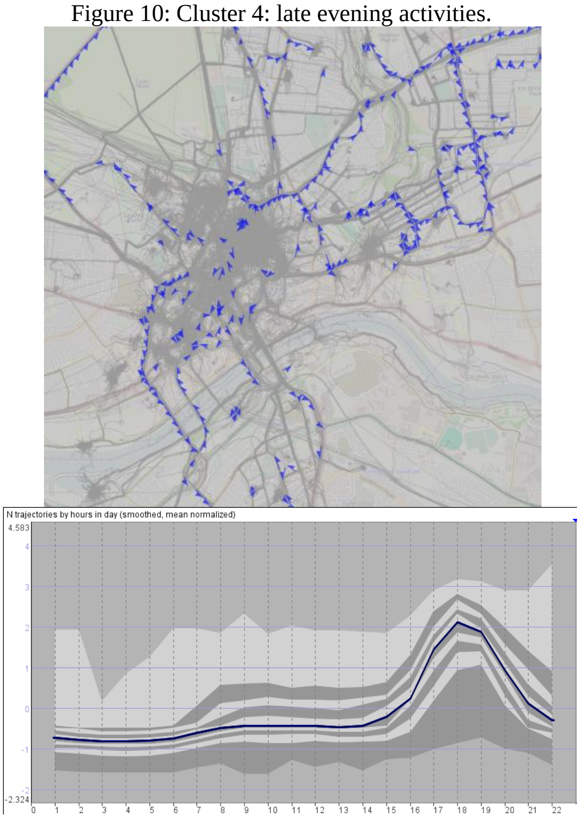


Figure 10: Cluster 4: late evening activities.



One can observe that different clusters are characterized by different temporal dynamics, see Figures 6-10. Hence, different roads have different temporal signatures, and, respectively, are used for different purposes.

4 Conclusions

Despite the large amount of the collected data, the high uncertainty of positions and small amount of data corresponding to street segments and individuals makes many types of potentially interesting analyses impossible.

References

[1] G.Andrienko, N.Andrienko, P.Bar, D.Keim, S.Wrobel. V *Visual Analytics of Movement*. Springer, 2013.

[2] N.Andrienko, G.Andrienko. Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Transactions on Visualization and Computer Graphics*, 2011, v.17 (2), pp.205-219.